



HHS PUBLIC ACCESS

Author manuscript

Neuroimage. Author manuscript; available in PMC 2017 March 06.

Published in final edited form as:

Neuroimage. 2016 November 01; 141: 399–407. doi:10.1016/j.neuroimage.2016.07.058.

Estimating functional brain networks by incorporating a modularity prior

Lishan Qiao^{a,b}, Han Zhang^b, Minjeong Kim^b, Shenghua Teng^b, Limei Zhang^{a,b}, and Dinggang Shen^{b,c,*}

^aSchool of Mathematics, Liaocheng University, Liaocheng 252000, China

^bDepartment of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA

^cDepartment of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea

Abstract

Functional brain network analysis has become one principled way of revealing informative organization architectures in healthy brains, and providing sensitive biomarkers for diagnosis of neurological disorders. Prior to any post hoc analysis, however, a natural issue is how to construct “ideal” brain networks given, for example, a set of functional magnetic resonance imaging (fMRI) time series associated with different brain regions. Although many methods have been developed, it is currently still an open field to estimate biologically meaningful and statistically robust brain networks due to our limited understanding of the human brain as well as complex noises in the observed data. Motivated by the fact that the brain is organized with *modular* structures, in this paper, we propose a novel functional brain network modeling scheme by encoding a *modularity prior* under a *matrix-regularized* network learning framework, and further formulate it as a sparse low-rank graph learning problem, which can be solved by an efficient optimization algorithm. Then, we apply the learned brain networks to identify patients with mild cognitive impairment (MCI) from normal controls. We achieved 89.01% classification accuracy even with a simple feature selection and classification pipeline, which significantly outperforms the conventional brain network construction methods. Moreover, we further explore brain network features that contributed to MCI identification, and discovered potential biomarkers for personalized diagnosis.

Keywords

Brain network; Functional magnetic resonance imaging (fMRI); Pearson’s correlation; Partial correlation; Sparse representation; Modularity; Low-rank representation; Mild cognitive impairment (MCI); Classification

*Corresponding author at: Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA. dgshen@med.unc.edu (D. Shen).

Introduction

Functional brain network (FBN) analysis has been successfully applied in both mining nontrivial topological structures in the healthy brain and revealing sensitive biomarkers for identifying neurological and psychological disorders (Stam, 2014; Fornito et al., 2015). However, different from the well-defined traditional networks (e.g., computer networks or social networks), the FBNs need to be estimated from data, prior to conducting any subsequent network analysis. It is generally challenging to construct high-quality FBNs because 1) we currently have limited understanding of the human brain, and 2) the observed data, e.g., the regional brain activity time series based on functional Magnetic Resonance Imaging (fMRI), tend to contain complex noises.

Despite of these challenges, many FBN estimation methods have been developed in the past few years (Smith et al., 2011), and most of them can essentially boil down to a *graph construction* problem. A graph $G(V, E)$ provides a mathematical tool to model a network, where V is a node (or vertex) set and E is an edge set that can be equivalently described by an edge weighting matrix W . In FBN, the nodes, according to different research focuses, may range from microscale of neurons to macroscale of brain regions. In this paper, we focus on the macroscopic FBN, where the nodes are defined as brain regions of interest (ROIs), and the edge between these ROIs can thus be determined by the relationship between their blood-oxygen-level dependent (BOLD) time series recorded by fMRI. For convenience of presentation, we will interchangeably use the terminologies (including ROIs, brain regions, and time series) to denote network/graph nodes, but do not distinguish their conceptual differences unless stated otherwise.

Pearson correlation (PC) coefficient is the most popular statistics for measuring the relationship between brain regions. However, PC only models the *full* correlations without excluding confounding effects from other brain regions. By contrast, *partial* correlation alleviates this problem by regressing out the potential influence from other brain regions. However, estimation of partial correlation involves inverting a covariance matrix, which may be ill-posed, especially when the sample size (i.e., the number of fMRI time points) is fewer than the dimensionality (i.e., the number of network nodes). To address this problem, a regularizer is usually introduced to the corresponding mathematical model (see Functional brain network construction section for more details). Such a regularization trick is *not only* for stabilizing the statistical estimation, *but also* for providing a principled way of incorporating priors into a network/graph learning framework (see Matrix-regularized network learning framework section for more details). In fact, many current FBN construction models can be interpreted in this framework. For examples, Huang et al. proposed to learn the brain connectivity by employing a sparsity prior (i.e., L_1 -norm regularizer) in the estimation of inversed covariance matrix (Huang et al., 2009); Lee et al. employed the same prior for the reconstruction of brain networks, based on the compressive sensing theory (Lee et al., 2011); Varoquaux et al. built FBN using a group sparsity prior ($L_{2,1}$ -norm regularizer) that constrains all subjects sharing the same network topology (Varoquaux et al., 2010); more recently, Wee et al. applied a similar prior based on a group LASSO regression for FBN construction (Wee et al., 2014).

However, the FBN commonly has more “structures” than just sparsity (Sporns, 2011). In this paper, inspired by the fact that the FBN is organized with a *modular* structure (Valencia et al., 2009), we present a novel FBN estimation scheme by encoding a *modularity* prior in form of matrix regularizer. Note that, in a recent work (Varoquaux et al., 2012), Varoquaux et al. proposed to describe the FBN by identifying its clique structure (similar to the modular structure here) in a *decomposable* graphical model. However, the clique structure in their methods was identified using a greedy approach, which may depend heavily on the *initial* graph and probably lead to local optima. In contrast, we formulate the FBN estimation as a *sparse low-rank graph learning* problem with a convex optimization model, and further propose an efficient algorithm to achieve its global optimal solution. Additionally, it is worth pointing out that the proposed method is not competing with Varoquaux et al.’s algorithm, since Varoquaux et al.’s algorithm can *in principle* work on any initially-constructed graph, including our estimated graph.

To verify the effectiveness of the proposed algorithm, we employ it in construction of FBN based on a real fMRI data set, and then use the estimated FBN to identify mild cognitive impairment (MCI) patients, which is important for early diagnosis and medical intervention of Alzheimer disease (AD). The experimental results show that the proposed method significantly outperforms the state-of-the-art methods. In particular, it achieves 89.01% classification accuracy even based on a simple feature selection (by means of *t*-tests with a fixed *p* value) and classification (via linear support vector machine (SVM) with default parameter $C=1$) pipeline. Also, we explore the selected features (i.e., network connections) in our method and found that most of the selected features tend to be biologically meaningful according to recent studies (Greicius, 2008; Albert et al., 2011). For facilitating efforts to replicate our results, we have shared both pre-processed data and codes in <http://www.nitrc.org/projects/modularbrain/>.

The rest of the paper is organized as follows. In the Functional brain network construction section, we review the related FBN estimation methods. In the Estimating brain network by incorporating modularity prior section, we present a matrix-regularized network learning framework, based on which we propose a novel FBN construction method including motivation, modeling, and algorithm. In Experiments section, we conduct experiments for corroborating the effectiveness of the proposed method. Finally, in the Conclusions section, we conclude our work with brief discussions.

Functional brain network construction

It is generally known that PC is currently the most widely used approach for estimating FBNs. According to a recent review (Smith et al., 2013), PC lies at the simplest extreme in a spectrum of FBN modeling methods, while dynamic casual modeling (DCM) is at the most complex extreme (Friston et al., 2003; Sengupta et al., 2016). Besides these two extremes, the popular FBN construction methods, from simple to complex, include partial correlation (Marrelec et al., 2006), regularized partial correlation (Friedman et al., 2008), Bayesian network (Ramsey et al., 2010), and structural equation modeling (McIntosh & Gonzalez-Lima, 1994), etc. Each of these methods, in our view, can be considered a *trade-off* between

PC and DCM by controlling the balance among the biological interpretability, computational efficiency, and statistical robustness.

In this paper, we mainly focus on the correlation-based methods, since they are successful in practice and have been empirically demonstrated to be more sensitive than the more complex or higher-order statistic-based methods (Smith et al., 2011). Below, we review two kinds of representative correlation-based methods in the Pearson's correlation and the Partial correlation and sparse representation sections.

Pearson's correlation

With pre-processing of the fMRI data (see Experiments section for details), we suppose that the brain has been parcellated into n ROIs, each of which corresponds to an observed time series $x_i \in \mathbb{R}^m$, $i = 1, 2, \dots, n$. Using the language of graph theory, we have a network node set $\{x_i\}_{i=1}^n$ in a m -dimensional space, and, without loss of generality, we rewrite the node set as an "ordered" matrix $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{m \times n}$. Then, our goal is to estimate the edge weight matrix $W \in \mathbb{R}^{n \times n}$ for FBN, given the data matrix X . The simplest way is to use PC coefficient, as defined below.

$$W_{ij}^{(PC)} = \frac{(x_i - \bar{x}_i)^T (x_j - \bar{x}_j)}{\sqrt{(x_i - \bar{x}_i)^T (x_i - \bar{x}_i)} \sqrt{(x_j - \bar{x}_j)^T (x_j - \bar{x}_j)}} \quad (1)$$

where $\bar{x}_i \in \mathbb{R}^m$ has all entries being the mean of the elements in x_i . Equivalently, $x_i - \bar{x}_i$ is a centralized counterpart of x_i . With the assumption that x_i is centralized by $x_i - \bar{x}_i$ and further normalized by $\sqrt{(x_i - \bar{x}_i)^T (x_i - \bar{x}_i)}$, PC can be simply expressed by $W_{ij}^{(PC)} = x_i^T x_j$, or, equivalently, $W^{(PC)} = X^T X$, which essentially corresponds to the estimation of covariance matrix Σ , with a multivariate normal distribution. Such a solution can also be achieved by the following regression problem (Lee et al., 2011).

$$\min_W \sum_{i,j=1}^n \|x_i - W_{ij} x_j\|^2 \quad (2)$$

To unify the related methods under a general *matrix-regularized* network learning framework (see Matrix-regularized network learning framework section for more details), we can equivalently formulate Eq. (2) to its matrix form as follows.

$$\min_W \|W - X^T X\|_F^2 \quad (3)$$

where $\|\cdot\|_F$ denotes the F -norm of a matrix.

Partial correlation and sparse representation

Despite its empirical effectiveness in modeling the FBN, PC, as mentioned in the Introduction section, can only detect *full* correlations, without excluding the confounding effect from other brain regions. By contrast, *partial* correlation is designed to handle this problem by regressing out the potential confounding variables. There are several strategies in statistics to calculate partial correlation. The widely used one is based on the estimation of inverse covariance matrix Σ^{-1} (a.k.a., precision matrix Θ). In particular, under the condition of normal distribution, we have that $\theta_{ij}(\Sigma^{-1})_{ij}=0$, if and only if the i -th and j -th variables are conditionally independent, and the partial correlation can then be defined by $-\theta_{ij} / \sqrt{\theta_{ii}\theta_{jj}}$ (Mardia et al., 1979).

However, the estimation of partial correlation can be ill-posed due to the singularity of the covariance matrix Σ , for example, when the sample size m (i.e., the number of fMRI time points in estimation of FBN) is fewer than the variable dimension n (the number of nodes in FBN). To address this problem, an L_1 -regularizer is generally introduced in the traditional estimation model, which results in two representative approaches. One is L_1 -regularized maximum likelihood estimation (Huang et al., 2009; Yuan & Lin, 2007), also known as graphical LASSO (Friedman et al., 2008); another is L_1 -regularized linear regression, which shares the same model with sparse representation (SR) and traditional LASSO (Meinshausen & Bühlmann, 2006; Peng et al., 2009). In this paper, we employ the latter (i.e., SR-based scheme) as one of the baselines for comparison with our proposed method, and its mathematical model can be obtained by the following objective function.

$$\min_W \sum_{i=1}^n \left(\|x_i - \sum_{j \neq i} W_{ij} x_j\|^2 + \lambda \sum_{j \neq i} |W_{ij}| \right) \quad (4)$$

For simplicity, we assume that all the node variables share the same regularized parameter λ as in (Meinshausen & Bühlmann, 2006). Similar to PC, we rewrite Eq. (4) to its corresponding matrix form as follows.

$$\min_W \|X - XW\|_F^2 + \lambda \|W\|_1 \text{ s. t. } W_{ii}=0, \forall i=1, \dots, n \quad (5)$$

where $\|\cdot\|_F$ and $\|\cdot\|_1$ are the F -norm and L_1 -norm of a matrix, respectively. The constraint $W_{ii}=0$ is equivalent to remove the variable x_i from X for avoiding trivial solutions. Note that the optimal solution W^* of Eq. (4) or Eq. (5) may be asymmetric. We empirically test and find that the asymmetry does not contribute to the final classification accuracy, and, thus, in our experiments we simply define the SR-based FBN as $W^{(SR)} = (W^* + W^{*T})/2$.

Estimating brain network by incorporating modularity prior

In this section, we first unify a class of FBN modeling methods under a general matrix-regularized network learning framework, which may benefit the understanding and comparison of the motivations behind different methods, and, more importantly, can provide a platform to develop the novel FBN learning method in this paper.

Matrix-regularized network learning framework

Here, we again first take PC and DCM as examples to illustrate what a “good” brain network model should be. In particular, PC fits the data *well* based on a statistical measure and works *efficiently* on modeling large-scale networks, but can only estimate simple linear relationship (implicitly based on Gaussian assumption), which lacks a reasonable biological meaning. On the other extreme, DCM is based on the *biological* mechanism, but can only model networks with relatively small scales (Smith et al., 2013). Naturally, we expect a good network model to achieve both perspectives. That is, it should 1) efficiently fit the data, and 2) effectively encode biological or physical prior knowledge. In fact, this trade-off between *data* and *knowledge* can be formulated by a regularized framework, which has been intensively studied in both statistics and machine learning fields. Here, we borrow the regularization concept for FBN construction, and present a more general framework by expressing priors (regularizers) in matrix form.

Based on the above discussion, we formulate a matrix-regularized network learning framework as follows.

$$\min_W f(X, W) + \lambda R(W) \text{ s. t. } W \in \Delta \quad (6)$$

where $f(X, W)$ is a data-fitting term, and $R(W)$ is a matrix-regularized term. Δ is a set including additional constraints on the network, such as non-negativity, symmetry, and positive semi-definiteness. It is worth pointing out that the data-fitting term plays a similar role to the loss function involved in some machine learning models, such as support vector machine and logistic regression, yet with different physical meanings. Specifically, the loss function tends to measure difference between the predicted values and ground truth, while $f(X, W)$ here measures which aspects of the data a network aims to capture. For example,

$f(X, W) = \|W - X^T X\|_F^2$ in PC aims at capturing the covariance structure in the data;

$f(X, W) = \|W - XW\|_F^2$ in sparse representation (or $f(X, W) = \text{tr}(SW) - \log |W|$ in graphical LASSO (Friedman et al., 2008)) aims at capturing the inverse covariance structure in the data based on a regression estimation (or maximum likelihood estimation).

For the regularized term $R(W)$, it is generally used in machine learning to prevent over-fitting. Here, such a term plays an important role in encoding biological or physical prior knowledge into the network construction. For example, sparsity is the most popular prior used in FBN construction, which is reasonable since it has been shown that the FBN is intrinsically sparse (Sporns, 2011). However, the FBN commonly has more “structures” than

just sparsity, such as small-worldness, scale-free topology, hierarchy and modularity (Sporns, 2011). We argue that some of these priors may guide to estimate a better FBN (see the toy example in the Model section for an intuitive explanation). In the Model section below, we attempt to encode the modularity prior into the FBN construction model, based on the matrix-regularized network learning framework, and further formulate it as a sparse low-rank graph learning problem.

The proposed method: Motivation, model, and algorithm

Motivation—Since Eq. (6) provides a regularized framework to model FBNs, now the main problem becomes how to encode the prior (i.e., modularity) in the form of matrix-regularizer. As we know, the modularity means that there exist some node groups (modules) in the network, and, within each group, the nodes are densely connected, while between these groups, the connections are sparse. As a result, the nodes within the same module tend to connect to each other with high probability, which may lead to some dependent rows/columns and then the low-rank edge weight matrix. In Fig. 1, we give a toy example to illustrate the relationship between the modularity of a network and the rank of its corresponding edge weight matrix.

In particular, the networks/graphs in the toy example include 7 nodes. Without loss of generality, we arrange the nodes in order, from 1 to 7, corresponding to the rows/columns of the edge weight matrix. Additionally, we assume that each node contains a link to itself, meaning that the diagonal elements of the edge weight matrix are all ones. As shown in Fig. 1, the graphs with strong modular structures, such as (c) and (d), generally have low rank. On the contrary, the graphs without clear modular structures, such as (b), (e) and (f), tend to be full rank. Especially for graph (e), it has the same sparsity as the graph (d), but the rank of the graph (d) is lower than (e), due to the modular structure in (d). Another special case is the complete or fully connected graph (a), in which any two nodes are linked by an edge, resulting in a rank-one matrix. Despite its low rank, such a graph is dense, not economic, and contains only a single “module”, which fails to provide any informative structure. Therefore, the modularity needs to be described by combining low-rank with sparsity. In other words, modularity is a *more powerful* constraint on the network structure, compared to sparsity, since modularity implies sparsity, not vice versa (see graph (e) in Fig. 1 for example). Now we can draw a preliminary conclusion that both sparsity and low-rank priors may contribute to estimate a reasonable brain network. Note, however, that this intuitive result is only illustrated by a toy problem with binary adjacent matrices. In the Experiments section, we will experimentally verify our model based on real data that 1) low-rank can help improve the network modularity structure, and 2) the sparse and low-rank combination works better than the model with only sparse (or low rank) regularizer in terms of classification.

Model—Based on the matrix-regularized network learning framework, we propose a novel functional brain network model by incorporating a modularity prior, as follows.

$$\min_W \|X - XW\|_F^2 + \lambda_1 \|W\|_0 + \lambda_2 \text{rank}(W) \quad (7)$$

Here, we use $\|X - XW\|_F^2$ as the data-fitting term, aiming to capture the inverse covariance structure in the data, due to its empirical effectiveness (Smith et al., 2011). According to the matrix-regularized framework, we can *in principle* use any reasonable options, such as

$\|W - X^T X\|_F^2$ involved in PC, to fit the data based on different motivations or applications. For the regularized terms, $\|W\|_0$ in Eq. (7) indicates the number of non-zero elements in W for measuring the sparsity of the network, and $\text{rank}(W)$ is the rank of matrix W , which cooperates with $\|W\|_0$ for modeling the modularity of the network. Unfortunately, the two regularizers are both non-convex with respect to W . Thus, we relax them to L_1 -norm $\|W\|_1$ and trace norm $\|W\|_*$ (a.k.a., nuclear norm), respectively, and obtain the following optimization model.

$$\min_W \|X - XW\|_F^2 + \lambda_1 \|W\|_1 + \lambda_2 \|W\|_* \quad (8)$$

where λ_1 and λ_2 are regularized parameters used to control the balance among the three terms in the objective function. Specially, when $\lambda_2=0$, Eq. (8) reduces to the network learning model based on the traditional sparse regression (Lee et al., 2011) given in Eq. (5); when $\lambda_1=0$, Eq. (8) reduces to the low-rank representation problem (Liu et al., 2010; Richard et al., 2012; Zhuang et al., 2012; Liu et al., 2013). In the Experiments section, we also include these two special cases for validating which prior (i.e., sparsity or low rank) contribute more to FBN construction and eventually MCI identification. Note that the proposed method shares a similar mathematical model with some machine learning algorithms (Liu et al., 2010; Richard et al., 2012; Zhuang et al., 2012; Liu et al., 2013), but has completely different explanation in both physical and statistical perspectives.¹ More specifically, Eq. (8) in machine learning, such as in (Zhuang et al., 2012), is used to capture the subspace structure in *samples* for clustering or semi-supervised learning, while, in our method, the model is used to capture the modular structure in *variables* for conducting correlation analysis. Furthermore, to our best knowledge, this is the first time that the technique is used to encode the modularity prior for estimating functional brain networks and identifying neurological disorders.

Algorithm—Although the objective function in Eq. (8) is convex, the L_1 and trace norm regularizers are both non-differentiable, which makes the problem nontrivial. Many algorithms have been developed in recent years to deal with this kind of problems (Richard et al., 2012; Zhuang et al., 2012). In this paper, we solve Eq. (8) based on *proximal method*

¹This is similar to the L_1 -regularized least square model, which is called LASSO for feature selection in statistics, but called sparse representation (SR) for signal recovery in signal processing field. In mathematics, LASSO and SR share the same model, yet with totally different meaning in both statistical and physical views.

(see (Combettes & Pesquet, 2011) and references therein) due to its simplicity and efficiency.

More specifically, according to the definition of *proximal operator* (e.g., Definition 10.1 in (Combettes & Pesquet, 2011)), the proximal operator of $\lambda_1 \|W\|_1$ is equivalent to the following soft thresholding operation on W ,

$$\text{prox}_{\lambda_1 \|\cdot\|_1}(W) = [\text{sgn}(W_{ij}) \cdot \max\{|W_{ij}| - \lambda_1, 0\}]_{n \times n} \quad (9)$$

Similarly, the proximal operator of $\lambda_2 \|W\|_*$ corresponds to a shrinkage operation on the singular values of W (Ji & Ye, 2009), as follows.

$$\text{prox}_{\lambda_2 \|\cdot\|_*}(W) = U \text{diag}(\max\{\sigma_1 - \lambda_2, 0\}, \dots, \max\{\sigma_n - \lambda_2, 0\}) V^T \quad (10)$$

where $U \text{diag}(\sigma_1, \dots, \sigma_n) V^T$ is the singular value decomposition (SVD) of matrix W .

Then, we consider the data-fitting term $f(X, W) = \|X - XW\|_F^2$ in Eq. (8). Since $f(X, W)$ is differentiable and its gradient (w.r.t. W) is $\nabla f(X, W) = -2X^T X + 2X^T XW$, we have the following gradient descent step,

$$W_k = W_{k-1} - \alpha_k \nabla f(X, W_{k-1}) \quad (11)$$

where α_k is the step size. To avoid the current W_k falling outside of the “feasible region” regularized by L_1 -norm $\|W\|_1$ and trace norm $\|W\|_*$, we orderly impose proximal operation on W_k by $\text{prox}_{\lambda_1 \|\cdot\|_1}(W_k)$ and $\text{prox}_{\lambda_2 \|\cdot\|_*}(W_k)$ given in Eqs. (9) and (10), respectively. Consequently, we have a simple algorithm for solving Eq. (8) as given in Table 1.

According to (Bertsekas, 2011), we can change the order of the two proximal operations involved in Steps 2 and 3 to update W , or simply conduct the min a random order, both of which can ensure convergence of the proposed algorithm. Additionally, in this paper, we employ the step size estimation strategy developed in (Ji & Ye, 2009) for achieving an optimal convergence rate. The source codes of this algorithm (with a cross validation procedure for parameter selection and performance evaluation) can be downloaded from <http://www.nitrc.org/projects/modularbrain/>.

In the next section, we apply the algorithm to construct FBNs, and then use the constructed functional connectivity as features for identifying MCI from normal aging.

Experiments

MCI is generally considered as a prodromal state of Alzheimer’s disease (AD). Thus, reliable detection of MCI is important for early medical intervention and may delay the

dementia. However, as an intermediate stage between normal aging and AD, MCI only involves subtle changes in functional connectivity, thus challenging effective identification. Therefore, MCI cohort provides a proper test bed to validate the proposed method.

Participants, data acquisition and pre-processing

The participants enrolled in this study were recruited via advertisements in local newspapers and media. All participants were right handed, with a normal or corrected-to-normal visual acuity, with no history of any neurological or psychiatric disorders, and without alcohol or drug abuse. Subjects with regular use of psychotropic, stimulants and beta-blockers were not included. Informed written consent was obtained from all participants, and experimental protocols were approved by local ethical committee. All recruited subjects underwent standard neuropsychological assessments, and were diagnosed by expert consensus panels.

Raw fMRI images were acquired using an echo-planar imaging sequence on a routine clinical whole-body 3 TMR scanner (TRIO, Siemens, Germany). The imaging parameters include: acquisition matrix size = 74×74 with 45 slices; voxel size = $2.97 \times 2.97 \times 3$ mm³; TE = 30 ms; TR = 3000 ms with 180 repetitions. During data acquisition, a simple CO₂ challenge, as in (Richiardi et al., 2015), was involved for another researching purpose on cerebrovascular reactivity (Richiardi et al., 2015; Cantin et al., 2011). Since the current study mainly focuses on classification rather than the biological meaning of “pure” resting-state networks, we conducted only the standard pre-processing for the acquired raw fMRI images, using the Statistical Parametric Mapping (SPM)² and DPARSFA (version 2.2) (Yan & Zang, 2010).

In particular, the first 10 fMRI images of each subject were discarded to allow signal stabilization. The remaining images were first corrected for different slice acquisition timing and head motion. The corrected images were registered to standard space, de-trended, and band-pass filtered (0.01–0.08 Hz) to remove the extremely low- and high-frequency artifacts, and further regressed out nuisance signals, including ventricle and white matter signals, as well as head motion parameters (Friston et al., 1996). Of note, the filtering also reduces the task-related block-shaped signals. Finally, a scrubbing operation was applied to the data for alleviating the impact of micro-head-movements on functional connectivity by removing the time points with frame-wise displacement larger than 0.5 (Yan et al., 2013). Since the scrubbing may remove large proportion of frames for some subjects, we exclude subjects with the remaining time points <80, thus leaving us 91 subjects³ (i.e., 45 MCIs and 46 NCs) with adequate data (i.e., >4 min) to model brain network. In Table 2, we list the main demographic information of the subjects included in this study.

Functional brain network construction

For each subject, we parcellate the above pre-processed BOLD signals into 90 ROIs based on the anatomical automatic labeling (AAL) atlas. Then, we use the mean signals on each

²<http://www.fil.ion.ucl.ac.uk/spm>.

³Here, we select the subjects with >80 time points mainly due to consideration of balance between valid time points and subjects. That is, there would be no sufficient subjects if only requiring sufficient time points. Also, we just use the first 80 time points of each subject to be consistent with each other, which actually provides an experimental condition for validating the FBN construction in small sample size cases.

ROI to estimate the FBN using different methods, e.g., PC, SR, and the proposed sparse low-rank (SLR) method. Note that, according to a recent survey (Smith et al., 2011), PC and SR are two of the most successful FBN construction methods. To investigate the contribution of different regularizers, we also include the low-rank (LR) method in our experiments by simply setting the regularized parameter $\lambda_1 = 0$ in Eq. (8). In Table 3, we list all the comparison methods under the matrix-regularized brain network learning framework.

In Fig. 2, we visualize the edge weight matrices⁴ of the FBNs constructed by four different methods. Here, we use the values of the regularization parameters as $\lambda=2^0$ for SR, 2^3 for LR, and $\lambda_1=2^{-1}$, $\lambda_2=2^3$ for SLR, respectively. We select these parametric values according to the best classification accuracies based on 1) all the network connections as features⁵ and 2) linear SVM (with $C=1$) as the classifier. For example, SR-constructed networks reach maximum accuracy when $\lambda=2^0$. From the results shown in Fig. 2, it can be observed that the combination of sparse and low-rank regularizers improves the network modularity, and the sparsity prior plays an important role in removing some weak (possibly noisy) connections.

Note that Fig. 2 just shows the edge weight matrices from a randomly selected subject. A natural problem is whether the FBNs from different subjects have similar structures (thus biologically more meaningful). Here, we simply measure this by the *mean* and the *standard deviation* of each edge weight across all subjects. Specifically, we define the mean edge

weight $\bar{W}_{ij} = \sum_{k=1}^{91} W_{ij}^{(k)} / 91$, and the standard deviation $S_{ij} = \sqrt{\sum_{k=1}^{91} (W_{ij}^{(k)} - \bar{W}_{ij})^2 / 90}$, where $W^{(k)}$ corresponds to the network of k -th subject constructed by SLR. In Fig. 3, we provide a visualization of the “mean network⁶” (left panel) and a statistics of the standard deviations of edge weights across all connections and all subjects (right panel). From Fig. 3, we can observe that the network structures tend to be consistent across subjects, since 1) the (positive) mean network preserves similar structures as the one shown in Fig. 2, and 2) most of the edges have relatively small standard deviations in connectivity strength across subjects.

Furthermore, for quantitatively evaluating the modularity, we employ Newman’s spectral algorithm (Newman, 2006) to calculate the modularity scores of differently constructed brain networks. Considering that the sparsity of a network may affect the modularity measures significantly, we report the results in Fig. 4 by sparsifying the networks based on different thresholds (from 0 to 0.99 with an increment of 0.1). Note that, since negative edge weights are invalid for Newman’s algorithm, we simply apply thresholds to the absolute values of the edge weights. That is, we remove the edge if its corresponding edge weight $|W_{ij}|$ is less than a certain threshold. Such an operation can not only remove some potentially noisy connections for improving the reliability of Newman’s algorithm, but also provide a platform to conduct fair comparison, especially for PC and LR, since their models do not include L_1 -regularized terms for controlling the sparsity. In addition, for avoiding the

⁴For convenience of comparison among different methods, all the weights are normalized to the interval $[-1, 1]$.

⁵Here, we use all the network connections as features since our goal is for visualization. In the Feature extraction/selection and classification section, we select the most discriminative features for classification and biological explanations.

⁶Since the positive and negative edge weights can offset each other, we only consider the positive edge weight when calculating the mean network for better interpretability.

randomness, we average the modularity scores on all the subjects for each method, with error bars denoting the standard deviation. It can be empirically observed from Fig. 4 that the combination of sparse and low-rank constraints can achieve better modularity (in terms of the best modularity score and the area under the curve). Note that the threshold operation here in fact plays a similar role as the L_1 -regularizer used for obtaining sparse networks.

Feature extraction/selection and classification

Once we obtain the functional brain networks (FBNs) for all subjects, the subsequent task is to classify MCI from NC according to the estimated FBNs. Then, the problem turns to determine which features should be used for classification. In practice, two kinds of strategies are often employed for FBN-based disorder identification. One is to extract features based on some graph measures, such as local clustering coefficients (Wee et al., 2014), and the other is directly using the network edge weights as features (Huang et al., 2009). Since different graph measures capture different aspects of network properties, it requires tricks or extra knowledge to design effective features. Therefore, in our experiments, we use the second strategy (i.e., using the raw edge weights as features), which can avoid the influence of differently extracted features on the validation of the FBN itself. Although the edge weight matrix theoretically includes all information of the network, it leads to the issue of high dimensionality. For example, an undirected network/graph with n nodes may generate $n(n-1)/2$ edges. In our study, $n=90$, and thus the feature dimensionality is 4005. To alleviate the curse of dimensionality, we first screen features by t -test with the empirically fixed p values⁷ before conducting classification using the linear SVM with default parameter (i.e., $C=1$). In our experiments, we do not employ any sophisticated feature extraction/selection and classifier design pipeline due to the two following considerations.

1. The main goal of these experiments is to verify the effectiveness of the proposed FBN estimation method. A complex feature extraction/selection and classification pipeline may confuse the validation of the network construction methods per se. For example, in a recent study (Wee et al., 2014), Wee et al. selected features sequentially using two filter-based approaches and one wrapper-based approach. We argue that it is hard to distinguish whether the feature selection strategies or network construction methods contribute to the ultimate classification accuracy.
2. It tends to over-fit the data if a complex classification pipeline is applied to limited training samples. In statistical learning theory, the sample size needs to increase exponentially with the model complexity (Bishop, 2006). For example, in our experiments, we have only 90 samples for training, even using leave-one-out (LOO) scheme. Besides the risk of overfitting, it is exceedingly challenging to select suitable values for the hyper-parameters involved in the feature selection and classification models, without sufficient training samples.

⁷In our experiments, we tested different p values from a candidate set $\{0.001, 0.005, 0.01, 0.05, 0.1\}$, and empirically found that 0.005 is the best value for PC, while 0.01 is the best value for SR, LR and SLR. Therefore, we simply fixed the $p = 0.005$ for PC, while 0.01 for other methods.

In Fig. 5, we show the basic procedure to conduct MCI vs. NC classification, which includes three main steps labeled by (1)–(3). In Step (1), we construct FBNs based on PC, SR, LR and SLR, respectively. Then, we conduct the LOO scheme for validating the effectiveness of different FBN construction methods. In particular, our dataset includes 91 subjects. In each LOO run, we use 90 samples (i.e., subjects or networks) for training and leave 1 sample for testing. The final performance is summarized by averaging the results from all 91 runs.

Since the regularized parameters involved in the FBN models may significantly affect the network structures and then the ultimate classification results (see Sensitivity to network model parameters section for more details), in Step (2) we select optimal parametric values by a grid search in a large range. For each regularized parameter (λ in SR, and λ_1, λ_2 in SLR), we use 11 candidate values in $[2^{-5}, 2^{-4}, \dots, 2^0, \dots, 2^4, 2^5]$. Note, however, that PC is parameter-free. For improving its flexibility and conducting fair comparison, we introduce a thresholding parameter in PC by preserving a proportion of strong edge weights. To be consistent with other methods, we also employ 11 candidate values $[100\%, 90\%, \dots, 10\%, 1\%]$ in our experiments. For example, 100% means all edges are preserved, and 90% means 10% weak edges are removed. Specifically, given a parameter (or parameter pair for SLR), we use 89 of the current training samples to select features (based on t -test with $p=0.005$ for PC and $p=0.01$ for other methods) and train a classifier (i.e., linear SVM with $C=1$). Then, we employ the remaining 1 sample to validate the classification performance. The optimal parameter (pair) is the one corresponding to the best validation performance (i.e., the average accuracy from the 90 *inner* LOO runs).

Note that the optimal network parameters (e.g., regularized parameters for SLR, and thresholding parameter for PC) may vary with different training sets. Therefore, in Step (3) we re-select features (also based on t -test) and re-train classifier (also linear SVM with $C=1$) based on the *current* training set with optimal network parameters. Finally, we classify the test sample using the selected features and trained classifier.

Classification results

In Table 4, we list the classification performance corresponding to 4 different FBN construction methods.

As shown in Table 4, the proposed FBN construction scheme with modularity prior achieves the best classification performance, outperforming even the result reported in a recent study (Wee et al., 2014) which used complex feature selection and classification pipeline. Here, we use only the simplest t -test and linear SVM for feature selection and classification, respectively. Therefore, we argue that the modularity prior, realized by combining sparse and low-rank regularizers, plays an important role in modeling and constructing functional brain networks. Also, we note that 1) PC involves less number of features because of using smaller p value of 0.005; 2) with the same p value of 0.01 for other methods, the networks estimated by SLR tend to include more discriminative features (i.e., edges); 3) the number of features used in SLR is relatively stable, since it has the smallest standard deviation across all LOO runs.

Sensitivity to network model parameters

For all methods, the ultimate classification accuracy is particularly sensitive to the network model parameters (i.e., thresholding values for PC, and regularized parameters for SR, LR and SLR). Therefore, in our above classification experiments, we conduct parameter selection in a large range by the inner LOO cross validation on an independent validation set separated from each training set. As an example, in Fig. 6 we show the classification accuracy corresponding to different parametric combination in the proposed method. Note that we simply compute the classification accuracy based on LOO test on all the subjects, since no parametric selection task is involved. It can be observed from Fig. 6 that 1) the ultimate results are extremely sensitive to the regularized parameters; 2) the modularity prior (i.e., combination of sparsity and low-rank priors) helps improve performance. In particular, we achieve the best accuracy (91.21%) with $\lambda_1=2^{-2}$ (for sparsity) and $\lambda_2=2^1$ (for low rank).

Besides the regularized parameters, the number of nodes in a network can also be considered as a free parameter. Therefore, we further conduct experiments for estimating and classifying FBNs with more nodes (including 200 ROIs (Craddock et al., 2012)). The experimental results show that SLR achieves a 72.53% accuracy, which also outperforms PC (68.13%), SR (51.65%) and LR (71.43%). Thus, we argue that the use of modularity (as a biologically inspired prior, or at least an assumption) can help learn better brain networks in terms of classification accuracy. Compared with the results on networks with 90 nodes, however, the performance of the methods (except PC) drops significantly. In our view, there are two main factors leading to this result. First, SR, LR and SLR all need to estimate the network by *inverting* a covariance matrix, as discussed previously. As we know, estimating an *inverse* covariance matrix with size of 200×200 is more challenging than that with size of 90×90 . In contrast, PC merely involves the estimation of a covariance matrix, and thus scales well. See the data-fitting terms in Table 3 for the difference between PC and other models. Second, with such a large-scale brain network, we need to select features from $200 \times (200 - 1) / 2 = 19,900$ edge weights, which is also more challenging based on the limited number of training samples (subjects). Therefore, for modeling larger-scale brain networks, e.g., voxel-based complex network (Zuo et al., 2012), the combination of covariance estimation (involved in PC) and modularity prior may be a feasible solution, and our proposed network learning framework in Eq. (6) provide a platform for doing this, which is an interesting research topic in the future.

Top discriminative features (network connections)

With the empirically optimal network parameters (i.e., $\lambda_1=2^{-2}$, $\lambda_2=2^1$) as shown in Fig. 6, we construct functional brain networks using the proposed method, and then employ *t*-test to sort features (i.e., network “connections”) according to their *p*-values. Consequently, we obtain 74 most discriminative connections with *p*-values < 0.01 , as shown in Fig. 7. Note that 1) the thickness of each arc in Fig. 7 represents the discriminating power of the corresponding connection (instead of its actual connectivity strength), which is inversely proportional to the corresponding *p*-value; 2) the color of each arc in Fig. 7 is randomly assigned just for better visualization. From such a set of connections, we find that several of them are biologically associated with MCI identification. Specifically, regions in the default mode network, such as the middle temporal gyrus, hippocampus, parahippocampus, superior

medial frontal gyrus, medial orbitofrontal gyrus, inferior parietal lobules, supramarginal gyrus and the precuneus, have strong discriminative ability, which is consistent with the previous neuroimaging biomarker reports and the pathology studies on MCI (Greicius, 2008; Albert et al., 2011).

Conclusions

Although various modeling methods have been developed in the past decades, it is still, to our best knowledge, an open problem to construct a reasonable functional brain network (FBN) from fMRI data. Thus, it is a potentially valuable topic to explore FBN construction based on different motivations, priors, or assumptions. In this paper, we propose to estimate FBN based on a matrix-regularized graph learning framework, in which the data-fitting term aims to capture the fMRI data distribution, while the regularized term encodes physical or biological prior into the network. More specifically, inspired by the modular structures of brain networks, we present a novel FBN construction method by incorporating modularity prior (via a combination of the sparse and low-rank regularizers) into the matrix-regularized graph learning framework. Finally, we use the constructed FBN for MCI identification, and achieve an encouraging accuracy (89.01%) even with simple feature selection and classification pipeline. It has been investigated that the brain networks exhibit more structures or properties than just the sparsity and modularity as discussed in this paper, such as assortativity, centrality, efficiency, hierarchy, synchronizability, hubs, rich club, small-worldness, and scale-free topology, etc. Additionally, we empirically found that the data-fitting term may affect the results significantly. Therefore, the next direction is to evaluate more data fitting terms and more priors under the proposed graph learning framework towards better human brain connectome.

Acknowledgments

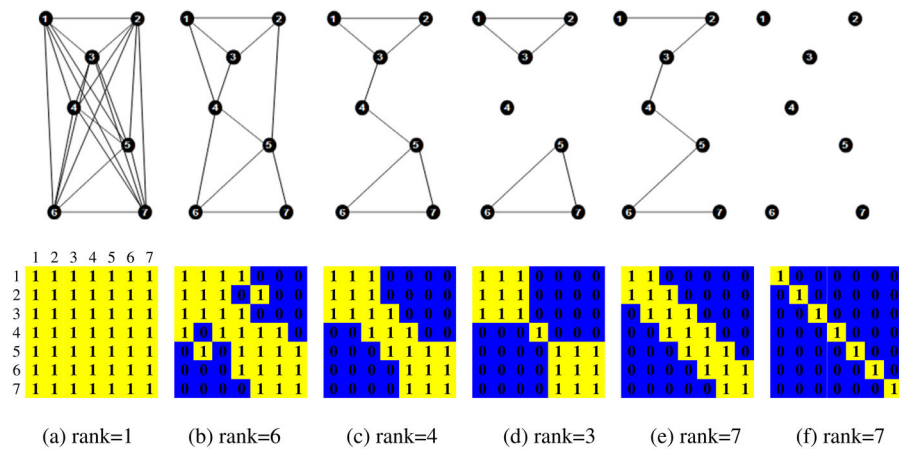
This work was partly supported by National Natural Science Foundation of China (61300154, 61402215), Natural Science Foundation of Shandong Province (2014ZRB019E0, 2014ZRB019VC), and NIH grants (AG041721, MH107815, EB006733, EB008374, EB009634).

References

- Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011; 7:270–279. [PubMed: 21514249]
- Bertsekas DP. Incremental gradient, subgradient, and proximal methods for convex optimization: a survey. *Optim Mach Learn*. 2011; 2010:1–38.
- Bishop, CM. *Pattern Recognition and Machine Learning*. Springer; New York: 2006.
- Cantin S, Villien M, Moreaud O, Tropres I, Keignart S, Chipon E, et al. Impaired cerebral vasoreactivity to CO₂ in Alzheimer's disease using BOLD fMRI. *NeuroImage*. 2011; 58:579–587. [PubMed: 21745581]
- Combettes, PL., Pesquet, J-C. *Fixed-point Algorithms for Inverse Problems in Science and Engineering*. Springer; 2011. Proximal splitting methods in signal processing; p. 185–212.
- Craddock RC, James GA, Holtzheimer PE, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum Brain Mapp*. 2012; 33:1914–1928. [PubMed: 21769991]

- Fornito A, Zalesky A, Breakspear M. The connectomics of brain disorders. *Nat Rev Neurosci*. 2015; 16:159–172. [PubMed: 25697159]
- Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*. 2008; 9:432–441. [PubMed: 18079126]
- Friston KJ, Harrison L, Penny W. Dynamic causal modelling. *NeuroImage*. 2003; 19:1273–1302. [PubMed: 12948688]
- Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R. Movement-related effects in fMRI time-series. *Magn Reson Med*. 1996; 35:346–355. [PubMed: 8699946]
- Greicius M. Resting-state functional connectivity in neuropsychiatric disorders. *Curr Opin Neurol*. 2008; 21:424–430. [PubMed: 18607202]
- Huang S, Li J, Sun L, Liu J, Wu T, Chen K, et al. Learning brain connectivity of Alzheimer's disease from neuroimaging data. *Adv Neural Inf Proces Syst*. 2009:808–816.
- Ji, S., Ye, J. An accelerated gradient method for trace norm minimization. *Proceedings of the 26th Annual International Conference on Machine Learning*; 2009. p. 457-464.
- Lee H, Lee DS, Kang H, Kim BN, Chung MK. Sparse brain network recovery under compressed sensing. *IEEE Trans Med Imaging*. 2011; 30:1154–1165. [PubMed: 21478072]
- Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y. Robust recovery of subspace structures by low-rank representation. *IEEE Trans Pattern Anal Mach Intell*. 2013; 35:171–184. [PubMed: 22487984]
- Liu, G., Lin, Z., Yu, Y. Robust subspace segmentation by low-rank representation. *International Conference on Machine Learning (ICML)*; 2010.
- Mardia, KV., Kent, JT., Bibby, JM. *Multivariate analysis*. Academic Press; 1979.
- Marrelec G, Krainik A, Duffau H, Péligrini-Issac M, Lehericy S, Doyon J, et al. Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage*. 2006; 32:228–237. [PubMed: 16777436]
- McIntosh A, Gonzalez-Lima F. Structural equation modeling and its application to network analysis in functional brain imaging. *Hum Brain Mapp*. 1994; 2:2–22.
- Meinshausen N, Bühlmann P. High-dimensional graphs and variable selection with the lasso. *Ann Stat*. 2006:1436–1462.
- Newman ME. Modularity and community structure in networks. *Proc Natl Acad Sci*. 2006; 103:8577–8582. [PubMed: 16723398]
- Peng J, Wang P, Zhou N, Zhu J. Partial correlation estimation by joint sparse regression models. *J Am Stat Assoc*. 2009:104.
- Ramsey JD, Hanson SJ, Hanson C, Halchenko YO, Poldrack RA, Glymour C. Six problems for causal inference from fMRI. *NeuroImage*. 2010; 49:1545–1558. [PubMed: 19747552]
- Richard, E., Savalle, P-A., Vayatis, N. Estimation of simultaneously sparse and low rank matrices. Presented at the *International Conference on Machine Learning (ICML)*; 2012.
- Richiardi J, Monsch AU, Haas T, Barkhof F, Van de Ville D, Radü EW, et al. Altered cerebrovascular reactivity velocity in mild cognitive impairment and Alzheimer's disease. *Neurobiol Aging*. 2015; 36:33–41. [PubMed: 25146454]
- Sengupta B, Friston KJ, Penny WD. Gradient-based MCMC samplers for dynamic causal modelling. *NeuroImage*. 2016; 125:1107–1118. [PubMed: 26213349]
- Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, et al. Network modelling methods for FMRI. *NeuroImage*. 2011; 54:875–891. [PubMed: 20817103]
- Smith SM, Vidaurre D, Beckmann CF, Glasser MF, Jenkinson M, Miller KL, et al. Functional connectomics from resting-state fMRI. *Trends Cogn Sci*. 2013; 17:666–682. [PubMed: 24238796]
- Sporns, O. *Networks of the Brain*. MIT Press; 2011.
- Stam CJ. Modern network science of neurological disorders. *Nat Rev Neurosci*. 2014; 15:683–695. [PubMed: 25186238]
- Valencia M, Pastor M, Fernández-Seara M, Artieda J, Martinerie J, Chavez M. Complex modular structure of large-scale brain networks. *Chaos: an Interdisciplinary. J Nonlinear Sci*. 2009; 19:023119.

- Varoquaux G, Gramfort A, Poline J-B, Thirion B. Brain covariance selection: better individual functional connectivity models using population prior. *Advances in Neural Information Processing Systems*. 2010:2334–2342.
- Varoquaux G, Gramfort A, Poline JB, Thirion B. Markov models for fMRI correlation structure: is brain functional connectivity small world, or decomposable into networks? *J Physiol Paris*. 2012; 106:212–221. [PubMed: 22326672]
- Wee CY, Yap PT, Zhang D, Wang L, Shen D. Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. *Brain Struct Funct*. 2014; 219:641–656. [PubMed: 23468090]
- Yan CG, Zang YF. DPARSF: a MATLAB toolbox for “pipeline” data analysis of resting-state fMRI. *Front Syst Neurosci*. 2010:4.
- Yan CG, Cheung B, Kelly C, Colcombe S, Craddock RC, Di Martino A, et al. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*. 2013; 76:183–201. [PubMed: 23499792]
- Yuan M, Lin Y. Model selection and estimation in the Gaussian graphical model. *Biometrika*. 2007; 94:19–35.
- Zhuang, L., Gao, H., Lin, Z., Ma, Y., Zhang, X., Yu, N. Non-negative low rank and sparse graph for semi-supervised learning. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2012. p. 2328-2335.
- Zuo XN, Ehmke R, Mennes M, Imperati D, Castellanos FX, Sporns O, et al. Network centrality in the human functional connectome. *Cereb Cortex*. 2012; 22:1862–1875. [PubMed: 21968567]

**Fig. 1.**

A toy example for illustrating the relationship between the modularity of a network and the rank of its corresponding edge weight matrix. Top panels: six networks with different sparsity and modularity; bottom panels: the edge weight matrices corresponding to the networks in the top panel. From the figure, it can be observed that the networks with modularity, such as (c) and (d), tend to be both sparse and low-rank.

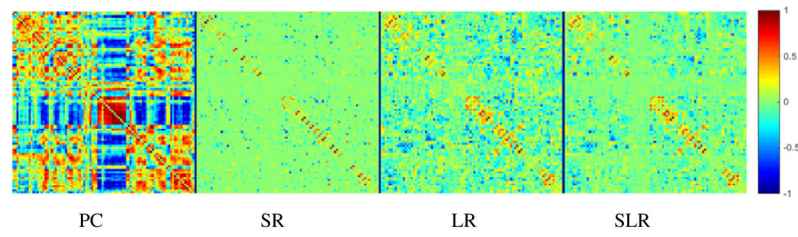


Fig. 2.

Four edge weight matrices of the same subject estimated by four different methods, i.e., PC, SR, LR and SLR. Compared with the first three methods, SLR can capture clearer modular structures by combining the sparse and low-rank regularizers.

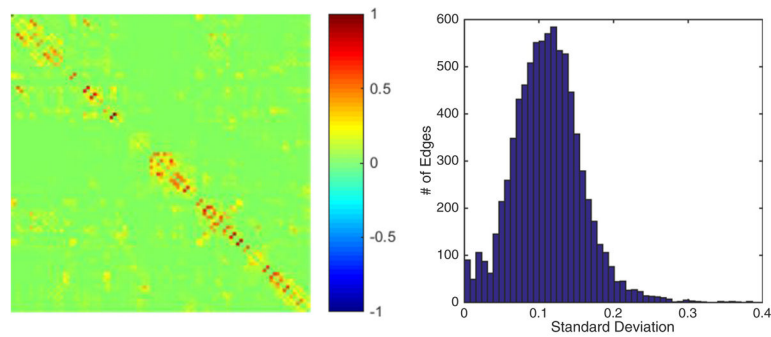


Fig. 3.

Left: The mean of positive edge weights across all subjects; right: The statistics of standard deviations of edge weights across different subjects.

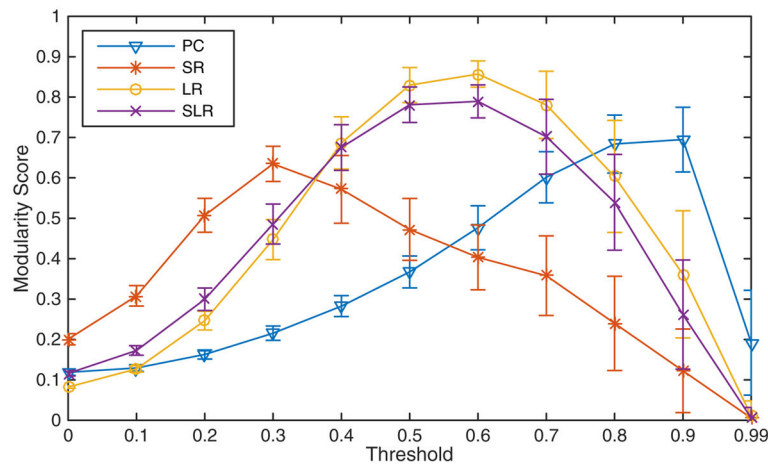


Fig. 4.

Modularity scores of networks constructed by different methods. All the networks are sparsified using different thresholds shown on the horizontal axis. Each modularity score is the average on all subjects, with a bar denoting the respective deviation. It can be easily observed that both sparsity and low-rank contribute to the modularity score.

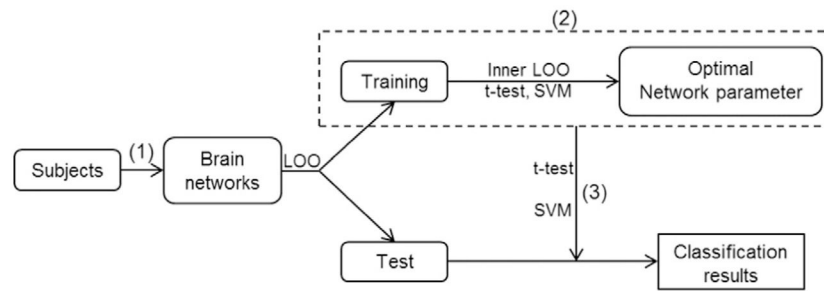


Fig. 5.
MCI identification pipeline used in our experiments.

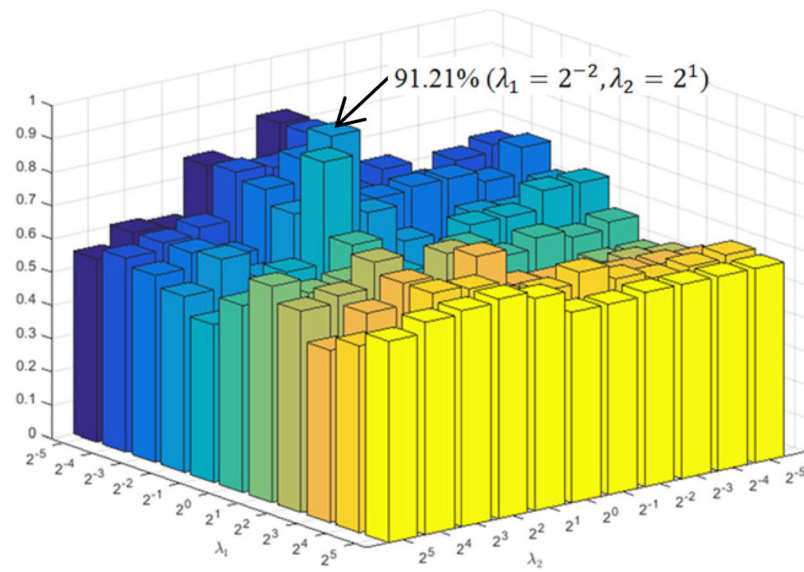


Fig. 6. Classification accuracy based on the networks estimated by the proposed method with different regularized parametric values in the interval $[2^{-5}, 2^5]$. The results are obtained by LOO test on all subjects.

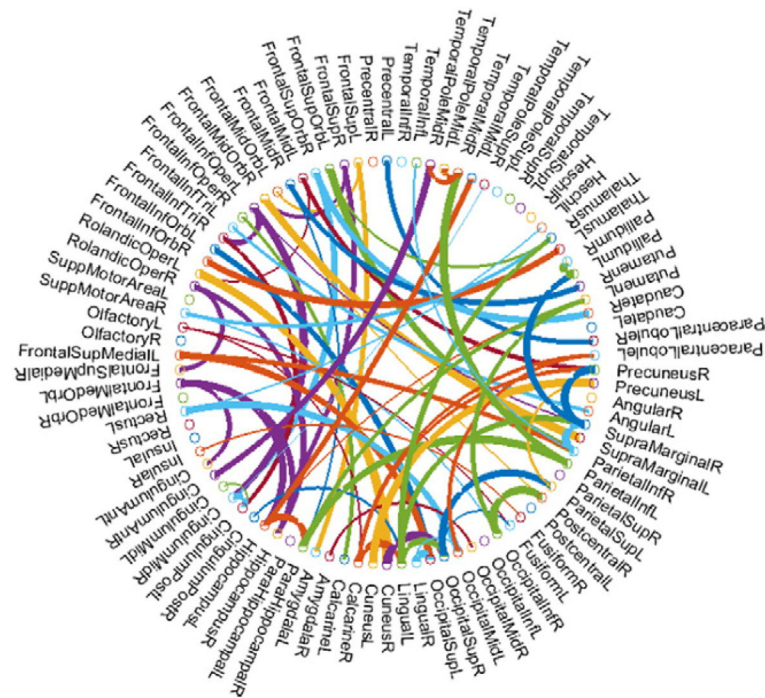


Fig. 7.

The most discriminative features (network connections) selected using t -test ($p < 0.01$). The thickness of each arc denotes the discriminating power of the corresponding connection.

This Figure is created by a Matlab function, circular Graph, shared by Paul Kassebaum.

<http://www.mathworks.com/matlabcentral/fileexchange/48576-circulargraph>.

Table 1

Algorithm for solving Eq. (8).

Initialize W	
Iterate:	
1	$W \leftarrow W - \alpha(-2X^T X + 2X^T X W)$
2	$W \leftarrow \text{prox}_{\lambda_1 \ \cdot\ _1}(W) = [\text{sgn}(W_{ij}) \cdot \max\{ W_{ij} - \lambda_1, 0\}]_{B \times B}$
3	$W \leftarrow \text{prox}_{\lambda_2 \ \cdot\ }(W) = \mathcal{U} \text{diag}(\max\{\sigma_1 - \lambda_2, 0\}, \dots, \max\{\sigma_B - \lambda_2, 0\}) V^T$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Demographic information of the used subjects.

	MCI	NC
# of subjects (male/female)	25/20	14/32
Age (mean \pm SD)	74.13 \pm 6.68	73.5 \pm 3.50
MMSE (mean \pm SD)	27.71 \pm 1.73	28.10 \pm 1.35

Table 3

Functional brain network construction methods, presented under the matrix-regularized network learning framework.

Method	Data-fitting term	Regularized term (prior)
Pearson's correlation (PC)	$\ W - X^T X\ _F^2$	N/A
Sparse representation (SR)	$\ X - XW\ _F^2$	$\lambda \ W\ _1$ (sparsity)
Low-rank representation (LR)	$\ X - XW\ _F^2$	$\lambda \ W\ _*$ (low-rank)
Sparse + low-rank rep. (SLR)	$\ X - XW\ _F^2$	$\lambda_1 \ W\ _1 + \lambda_2 \ W\ _*$ (modularity)

Table 4

Comparison on classification performance and the number of involved features for 4 different methods. The last column shows the mean \pm standard deviation of the feature numbers used in all 91 LOO runs. The bold numbers indicate the best results for the accuracy, sensitivity and specificity, respectively.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	#Features
Pearson's correlation (PC)	69.23	71.11	67.39	45.1 \pm 8.0
Sparse representation (SR)	71.43	71.11	71.74	60.4 \pm 7.9
Low-rank (LR)	79.12	80.00	78.26	62.5 \pm 4.0
Sparse low-rank (SLR)	89.01	86.67	91.30	72.4 \pm 3.3